

Classification and Prediction Students' Literacy Ability Using Machine Learning

by Heri Retnawati, Mathilda Susanti, Limala Ratni S. Kharismawati, Purwoko Haryadi Santoso, Kartianom

ABSTRACT

Literacy is an important ability and skill that for students in the 21st century because it is closely related to problem-solving abilities in various contexts and in everyday life. Literacy predictions using various variables and their classification need to be done to provide input on education policy, but based on existing research, this has not been conducted much by other researchers, including using machine learning. In this regard, this research aims to carry out predictive analysis and classification of students' language literacy in Indonesia using machine learning.

This is an exploratory descriptive research with a quantitative approach. Research data was collected using documentation data from the SEAQIL Indonesian Language Institute. The variables involved in this research are literacy achievement, personal attitudes, language use, parental support, school facilities, library use, and students' reading habits which were collected from various instruments. The validity of the instrument was analyzed using confirmatory factor analysis, and reliability was estimated with the conditional reliability coefficient of the IRT. Data analysis was carried out using machine learning assisted by R software, by dividing the data into two parts, namely 80% training data and 20% validation data.

Based on the research results, it can be concluded that in predictions using the LR model, there are five variables related to student factors, three variables related to school factors, and two variables related to provincial factors. It can be seen that the LR model is dominated by variables related to student factors in explaining student literacy achievements, but the level of importance or predictive power is not as high as variables from school and provincial factors. On the other hand, the SVM, KNN, and RF models actually identify three variables related to student factors, five variables related to school factors, and two variables related to provincial factors. These three models show that variables related to school and province are crucial factors in explaining student literacy achievements. In classification, the research results show that non-cognitive variables such as school quality and students' regional origin combined with students' grades in language subjects are the most important features in predicting Indonesian students' literacy abilities. Non-cognitive variables are able to produce better prediction performance than if we combine them with the results of cognitive assessments in the model. The implications at theoretical and practical levels of the findings are discussed in this article.

Kata Kunci: *Literacy, clasification and prediction, machine learnig*